Requested Patent: EP0676882A2

Title:

SPEECH RECOGNITION SYSTEM WITH DISPLAY FOR USER'S CONFIRMATION. ;

Abstracted Patent: EP0676882 ;

Publication Date: 1995-10-11 ;

Inventor(s): HAIMI-COHEN RAZIEL (US); REED ADAM VICTOR (US) ;

Applicant(s): AT _T CORP (US) ;

Application Number: EP19950302109 19950329 ;

Priority Number(s): US19940223810 19940406 ;

IPC Classification: H04M1/27 ; H04M3/50 ;

Equivalents: CA2143980, CN1115931, JP7288588

ABSTRACT:

A speech recognizer system for use with a telecommunication network wherein an input signal generated onto the network from a first terminal is directed to a speech recognizer for estimating the verbal content of the input signal. The speech recognizer or associated equipment then directs an estimate of the verbal content as an output signal back to the first terminal, the estimate including one or more approximations of the verbal content of the input signal. At the first terminal the user then confirms a correct estimate, or selects from a plurality of approximations, the verbal content of the input signal.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number : **0 676 882 A2**

(12) EUROPEAN PATENT APPLICATION

(21) Application number : **95302109.4**

(51) Int. Cl.[6] : **H04M 1/27,** H04M 3/50

(22) Date of filing : **29.03.95**

(30) Priority : **06.04.94 US 223810**

(43) Date of publication of application :
11.10.95 Bulletin 95/41

(84) Designated Contracting States :
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

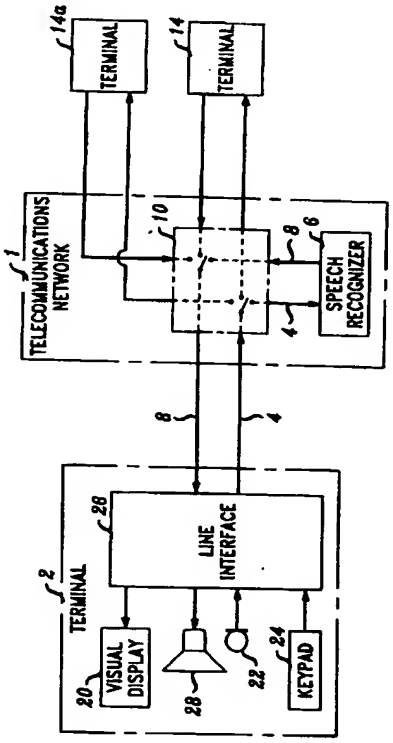
(71) Applicant : **AT & T Corp.**
32 Avenue of the Americas
New York, NY 10013-2412 (US)

(72) Inventor : **Haimi-Cohen, Raziel**
30 N. Derby Road,
Springfield
New Jersey 07081 (US)
Inventor : **Reed, Adam Victor**
23 Longfellow Terrace,
Morganville
New Jersey 07751 (US)

(74) Representative : **Watts, Christopher Malcolm Kelway, Dr. et al**
AT&T (UK) Ltd.
5, Mornington Road
Woodford Green Essex, IG8 0TU (GB)

(54) Speech recognition system with display for user's confirmation.

(57) A speech recognizer system for use with a telecommunication network wherein an input signal generated onto the network from a first terminal is directed to a speech recognizer for estimating the verbal content of the input signal. The speech recognizer or associated equipment then directs an estimate of the verbal content as an output signal back to the first terminal, the estimate including one or more approximations of the verbal content of the input signal. At the first terminal the user then confirms a correct estimate, or selects from a plurality of approximations, the verbal content of the input signal.



EP 0 676 882 A2

Field of the Invention

The present invention relates to the field of voice or speech recognition and more specifically to speech recognition in a network.

Background of the Invention

Speech recognition is the process by which an audio input signal is received and the verbal content of the input signal is determined. The verbal content is then further processed to obtain the desired action. The verbal content can be a transcription of the speech of the input or merely a general statement of content. Additionally, the speech recognizer itself can be located at the user terminal, as in United States Patent No. 5,111,501 or in the network as in United States Patent No. 4,922,519.

Since error exists in the determination of verbal content, systems have been established whereby the user is asked to repeat the request if the speech recognizer is unsure of the content. In this regard, a study of customer interaction with speech recognizers is reported in "Serving Customers With Automatic Speech Recognition - Human-Factors Issues" by Wattenbarger et al, AT&T Tech. J., May/June 1993, pp. 28-41.

Summary of the Invention

The present invention is directed to a speech recognizer system for use with a network comprising a first terminal from which an input signal is generated onto the network, speech recognizer means for estimating verbal content of the input signal, feedback means for directing an output signal comprising one or more approximations to the first terminal and confirmation means including selection means at the first terminal for confirming the correct approximation.

Preferably, the speech recognition means creates output signals which are digital to increase transmission speed. It is also preferred that the first terminal have a visual display on which to display the estimate, either one approximation at a time, all at once or a number therebetween. Of course, an audio feedback of the estimate may be preferred in such situations as a car phone where the user cannot easily and safely view a visual display or at a terminal without a visual display.

The present invention further includes a method for speech recognition including the steps comprising placing an input signal onto a network, estimating the verbal content of the input signal on the network and transmitting the estimate back to the first terminal for confirmation. Of course, if the speech recognizer is certain of the speech content from the input signal, feedback of an estimate to the first terminal need not be performed.

This invention also provides for error reduction in speech recognition systems comprising the steps of providing an estimate of the verbal content of an input signal to the user and receiving confirmation of a correct approximation from the user.

Brief Description of the Drawings

The figure is a schematic block diagram of the system of the present invention.

Description of the Preferred Embodiment

In the preferred embodiment, the present invention is used with a telecommunications network 1 connected to a first terminal 2 from which a user generates an input signal 4 onto the network 1. The input signal 4 is transmitted to a speech recognizer 6, also within the network 1, where an estimate, including one or more approximations of the verbal content of the input signal 4, is made. The estimate of the verbal content is converted to an output signal 8 which is transmitted back to the first terminal 2 for user confirmation. Once the correct verbal approximation is confirmed by the user at the first terminal 2, the information is processed to complete the exchange, for example as described in United States Patent No. 4,922,519 to Daudelin.

In its preferred embodiment, the first terminal 2 is an automatic device or a user operated device such as a telephone plugged into the network having a visual display 20 such as Caller I.D. In practice, however, it is understood that any device producing a variable signal on the network can be used.

Preferably, a user is able to speak into the microphone 22 of a telephone, for example, to recite a desired number to be dialed, to recite whether a call is to be collect, charged to a calling card, person-to-person, etc., to recite a person's name for which a number is requested or to provide other information. The speech is generated onto the network as an input signal 4, either in analog or digital format depending on the equipment making up the first terminal line interface 26.

Within the network 1 is a speech recognizer 6 which receives and processes the input signal 4. A suitable speech recognizer is a CONVERSANT CVIS, manufactured by AT&T.

If the speech recognizer 6 is sure of the verbal content of the input signal 4 (greater than a predetermined percent certain, e.g. >90% certain) the speech recognizer 6 passes the information to the switch 10 or other processing device as required. If, however, the speech recognizer 6 is not sure of the verbal content of the input signal 4 it will provide an estimate of the verbal content in an output signal 8, comprising one or more approximations of the verbal content, to the first terminal 2 for confirmation of the estimate or

selection of an approximation.

Preferably, the first terminal 2 has a visual display 20 and the output signal 8 is in a digital format when sent to the first terminal 2 to speed the movement of the output signal 8 on the network to the first terminal 2. Preferably the terminal will have a modem to decode the digital signal for presentation. Alternatively, the output signal 8 will be in DTMF, which can be used with most all current terminal system, that is decoded and presented in visual form at the terminal. The display 20 of the terminal 2 will then decode the signal 8 and visually present it to the user.

At the first terminal 2 the output signal 8 is received and at least a portion thereof, i.e. one approximation, is presented to the user for confirmation. In its preferred embodiment the most probable approximation will be displayed first, followed by the next most probable if the first is not selected or confirmed.

Of course, although a visual display 20 is preferred, the presentation can be audio via a speaker 28, visual or both, with visual alone or in combination with audio being preferred in most instances due to the speed of presentation and reduced need for user alertness where the visual information can stay on a display 20 until the user wishes to remove it, by confirming the correctness or selecting the next approximation.

However, for car phones where the user has his eyes on the road or with telephones that do not have a visual display, audio presentation via speaker 28 is available alone or along with the visual display 20.

In situations where an audio presentation is made, "barge-in" capability is especially important so a user does not need to wait for the end of the audio presentation to confirm a correct approximation or request the next selection. The barge-in feature allows the user to make a confirmation or request another approximation, by depressing a key on the keypad 24 or speaking into the microphone 22 during the presentation, thereby terminating the presentation of the previous approximation without having to listen to the entire presentation.

The preferred visual display 20 can be of any type, including a Caller I.D. where a line or more of alphanumeric text is presented in an LCD display, a P.C. monitor, a CRT display, a vacuum fluorescent display, an LED display, a video telephone, a still image telephone, etc.

In implementing the speech recognition system of the present invention, a communication protocol must be defined for transmitting the output signal 8 from the network 1 to the first terminal 2. Definition of the protocol requires that the variety of possible terminal types and visual displays present in the network be taken into account. Several methods are currently envisioned herein, including a bidirectional protocol, a terminal specific protocol and a unidirectional protocol.

A bidirectional protocol requires that the terminal 2 respond to the network prompt and describe the capabilities of the terminal 2. The network can then direct an output signal 8 to the terminal 2 which matches the capabilities of the terminal 2. For example, if the line interface 26 of the first terminal 2 has a high speed modem, the output 8 will be set faster using the modem protocol. If the terminal 2 is a videophone or still image telephone, the system will generate an output signal 8 comprising a video image for transmission to the visual display 20 terminal 2. If the terminal 2 can display more than one approximation, the estimate may be transmitted by the network for visual display of more than one approximation and prompt the user by synthesized speech, etc., to choose. In the bidirectional protocol a terminal which does not respond to the prompt will be considered to not have any visual display 20 and the output 8 will be in the form of synthesized speech.

With a terminal specific protocol, the network 1 stores a table of the identities of each terminal 2 and utilizes a terminal specific protocol based on the information on the specific terminal. This approach, however, would only be effective in a small network where a network administrator has control over all of the installed terminals.

With a unidirectional protocol the network transmits both a digital feedback for visual display and an audio synthesized speech feedback for audio presentation to the user at the first terminal 2. The format is fixed and the specific terminal can ignore or display the digital feedback for visual display. Of course, this is the most simple protocol, however, it does not allow for customization to specific terminals.

When the presentation is made to the user at the first terminal 2, the user is able to confirm a correct estimate. This includes the ability to indicate that a correct estimate is displayed or request another alternative if additional alternatives are available. If a multi-line display, e.g. a CRT display is used, the confirmation means includes selection means to select from the approximations displayed, to scroll down or bring up a new screen of additional approximations, etc. Such means includes a keypad 24 ora microphone 22 for voice input.

Additionally, the feedback can be augmented with other information resulting from the query of a database with a recognized input or the most closely matching approximations of an input. For example, in a telephone directory application response to an input signal may include an estimate including the most closely matching name or names together with the corresponding telephone numbers. Similarly, in an exchange request the cost of calling each of the approximations can be included. In such applications the confirmation feature can include automatic dialing of the selected approximation or a request for the next screenful of approximations.

Example

A user at a first terminal 2, having a visual display 20 comprising a Caller I.D. device, associated with a network 1 which employs a speech recognition facility lifts a handset and dictates a telephone number of another terminal 14 he wishes to be connected to into the microphone 22. An input signal 4 is placed onto the network 1 and is directed to a speech recognizer 6. The speech recognizer 6 estimates the verbal content of the input signal 4, i.e. the telephone number recited by the user, producing an estimate of the verbal content of the input 4 comprising three (3) approximations.

The estimate is coded in a digital format and is transmitted as an output signal 8 in a DTMF format to the first terminal 2. The approximation considered most probable by the speech recognizer is displayed on the Caller I.D. LCD display 20 and a speech synthesized voice recites the number through the speaker 28. If the displayed approximation is the one the user desires, the user depresses the "*" key on the keypad 24 to confirm the number, thereby directing the network to attempt to reach the other terminal 14 associated with that number. A barge-in feature stops the speech synthesized voice when the key is depressed.

If the first approximation displayed is not the correct number, the user depresses the "#" key on the keypad 24 and the next most probable approximation appears. Again, a barge-in feature stops the synthesized voice reciting the first approximation and begins reciting the next approximation when the "#" key is depressed on the keypad 24. When the correct approximation is displayed the user confirms the approximation by depressing the "*" key on the keypad 24 and the call to the other exchange 14 is placed on the network 1.

While the present invention has been described in detail with reference to specific embodiments thereof, it will be apparent to those skilled in the art that various changes and modifications can be made without departing from the scope of the invention.

Claims

1. A speech recognizer system for use with a telecommunication network comprising first terminal means from which a user generates an input signal onto the network, speech recognizer means in the network for providing an estimate of the content of the input signal, feedback means for directing an output signal of the estimate from the speech recognizer means to the first terminal means, said estimate comprising more than one approximation of the content of the input signal, and confirmation means at the first terminal means for confirming or selecting a correct approximation of the content of the input signal.

2. The speech recognizer system of claim 1 wherein the output signal comprises a digital signal.

3. The speech recognizer system of claim 2 wherein the first terminal means comprises a modem and visual display means for presenting at least one of the more than one approximation of the estimate of the content of the input signal to the user at one time.

4. The speech recognizer system of claim 3 wherein the visual display means is a caller I.D. device, an LCD display, a CRT display, a P.C. monitor, a video terminal display, a video telephone, a still image video display, a still image telephone, an LED display or a vacuum fluorescent display.

5. The speech recognizer system of claim 1 wherein the output signal comprises synthesized speech in analogue form.

6. The speech recognizer system of any of the preceding claims wherein the output signal further comprises additional information related to the content of the input signal.

7. The speech recognizer system of claim 6 wherein the content of the input signal comprises an identification of second terminal means and the output signal comprises telephone numbers corresponding thereto.

8. A method of speech recognition in a telecommunication network comprising the steps of estimating within the network the content of an input signal placed onto the network from first terminal means, and transmitting an output comprising an estimate of the content of the input signal back to the first terminal means for confirmation, said estimate comprising more than one approximation of the input signal.

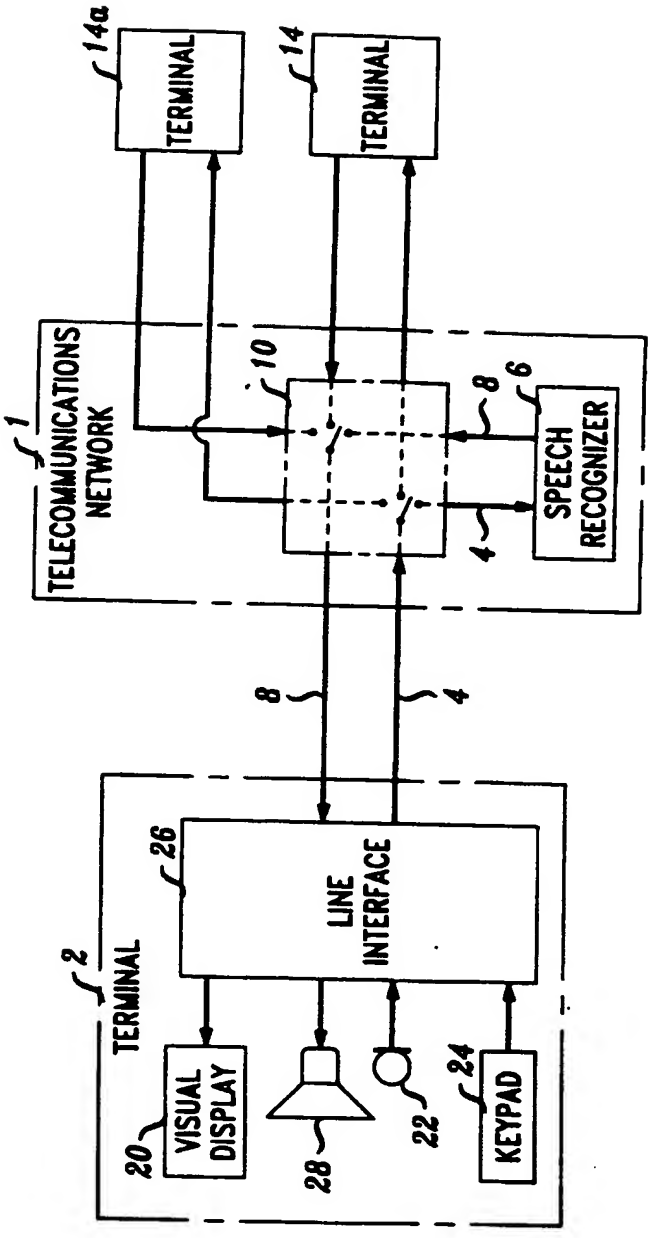
9. The method of claim 8 further comprising the step of digitizing the output signal prior to the transmission to the first terminal means.

10. The method of claim 8 or claim 9 further comprising visually displaying one or more of the approximations of the content of the input signal at the first terminal means.

11. The method of any of claims 8 to 10 wherein the approximations are presented sequentially at the first terminal means with a most probable approximation presented first and a successive next

most probable alternative presented after said most probable approximation only if the most probable approximation has not been confirmed.

12. A method of error reduction in speech recognition in a telecommunications network comprising the steps of providing an estimate of the content of an input signal placed on the network from a user at first terminal means back to the first terminal means, said estimate comprising more than one approximation of the content of the input signal, and providing confirmation or selection of a correct approximation from the first terminal means onto the network.


14a
TERMINAL
14
TERMINAL
1
TELECOMMUNICATIONS NETWORK
10
8
6
SPEECH RECOGNIZER
4
8
4
2
TERMINAL
26
LINE INTERFACE
20
VISUAL DISPLAY
28
22
24
KEYPAD